Fundamentals of Decision Units and Sampling Design

nvironmental Monitoring and Data Quality Worksho
30 March – 3 April 2009

Thomas Georgian, Ph.D.
US Army Corps of Engineers
Environmental and Munitions
Center of Expertise

Target Objective

- The best we can reasonably hope for is to be slightly less confused.

Overview

- Define “sampling unit” (SU).
- Define “decision unit” (DU).
- Discuss some fundamental statistical concepts.
- Address some common misconceptions.
- Discuss what constitutes a “hot spot” vs. a DU.

Sampling Unit (SU)

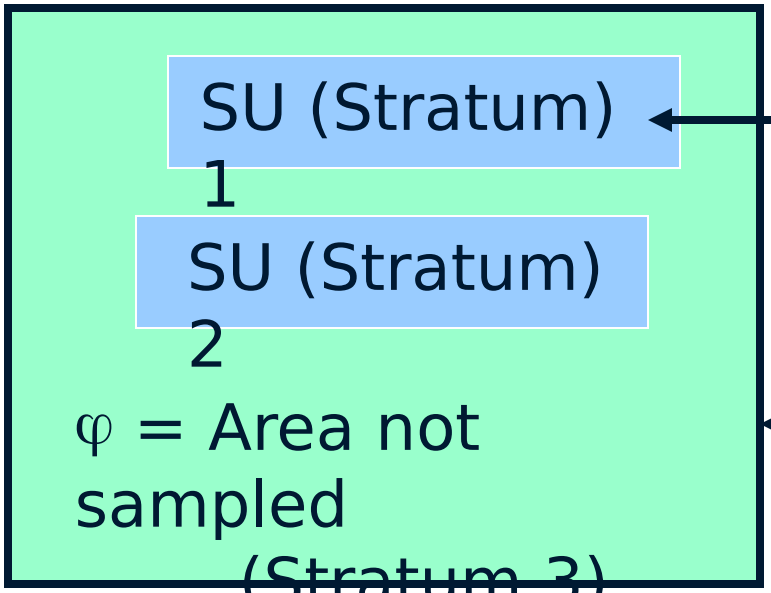
- **What will be sampled** (e.g., portion of site, a study area or pre-defined volume of material represented by a sample)

1. Must have well defined boundaries.

1. There is a separate pre-defined sampling design that produces representative results.

Decision Unit (DU)

- Unit with well defined boundaries consisting of one or more **Sampling Units** or **strata** for which a decision will be made.

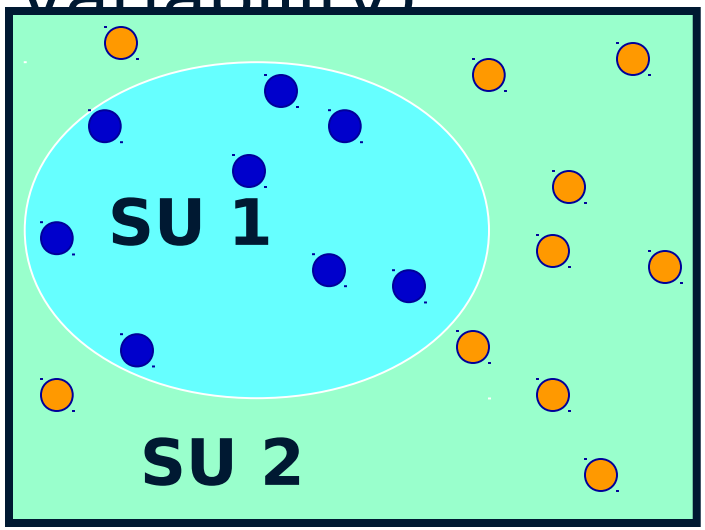


Stratum: Portion (subset) of a DU that may or may not be sampled. (If sampled stratum is a SU.)

DU = SU 1 + SU 2 + φ

Stratification

- Dividing the DU into strata, typically various SUs (e.g., using historical information, prior data or knowledge of "hot spots") to optimize the sampling design (e.g., to decrease variability)



DU = SU 1 + SU 2

Stratified Random Sampling

SU 1: Stained area

SU 2: Remaining portion of DU

$\sigma_{DU} > \sigma_{SU1}, \sigma_{SU2}$

Decision Unit (DU)

- Typically

1. Multi Increment® (MI) sampling is done for the SUs (i.e., composite sampling of solids consistent with Pierre Gy's theory)

2. Project Data Quality Objectives (DQOs) are established for mean analyte concentrations for the SUs to make decision about the DU.

DU & SUs Defined Using 7-Step DQO Process (QA/G-4)

1. State problem
2. Identify decision
3. Identify decision inputs
4. **Define study boundaries – DU & SUs defined**
5. Develop decision rule
6. Specify tolerances for decision errors
7. Optimize study design

Determining DU & SUs (Grabs and Composites)

- Established after steps 1 – 3 of DQO Process.
  - Cannot be defined until end use of data is known.
  - SU = smallest unit of interest for any end use.
- Developed from Conceptual Site Model (CSM).
  - Depends on nature of contaminant (e.g., mobile vs. immobile)
  - Nature of receptors (sizes of exposure units)

Example of how NOT to define a SU from actual Site Investigation (SI)

- DU is 640-acre aerial gunnery training range used in 1942 – 1945. Ground targets "engaged" with aircraft mounted machine guns.
- SU (center of target area) is 100 m x 100 m (2.5 acres) to a depth of 5 cm.
  - Triplicate MI samples of 100 increments collected.
  - Tested for explosives by 8330B (e.g., NG & DNT from primers & propellants of 0.30 & 0.50 caliber small arms).

SU must be consistent with DQOs

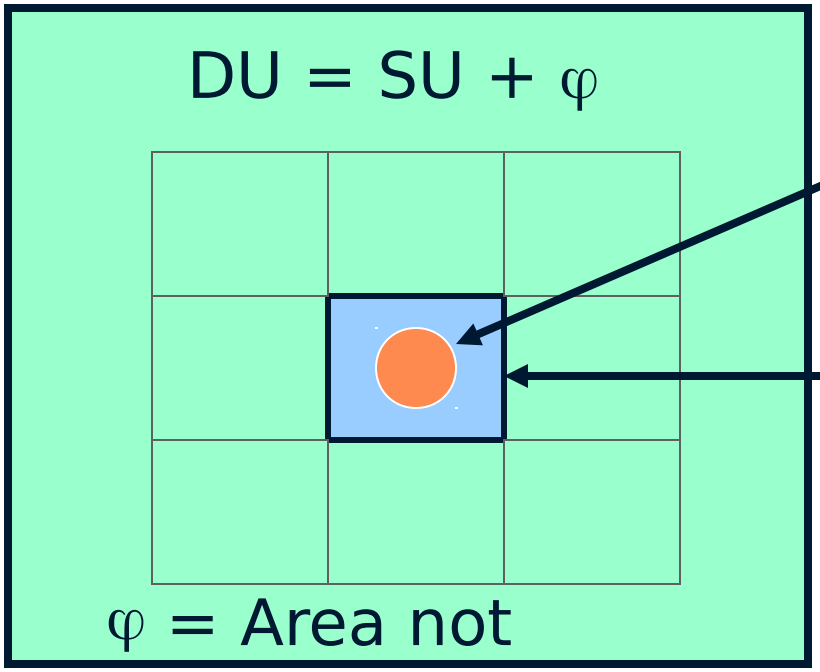
- DQOs - Excerpt from SI Report:

  *Conducting a human health screening level risk assessment (SLRA) for soil by comparing soil sample analytical results to the more conservative of Texas Risk Reduction Program (TRRP) Tier 1 residential protective concentration limits (PCL) for **0.5-acre sites** and USEPA Region 6 residential medium-specific screening levels (MSSL)…*

- SU (2.5 acres) = 5 × exposure units (0.5 acres).
  Results will be biased low owing to "dilution."

  Conclude $\mu_{0.5} < L$ if $\mu_{2.5} < L/5$, L = Action Level

Alternative Approach



Target center

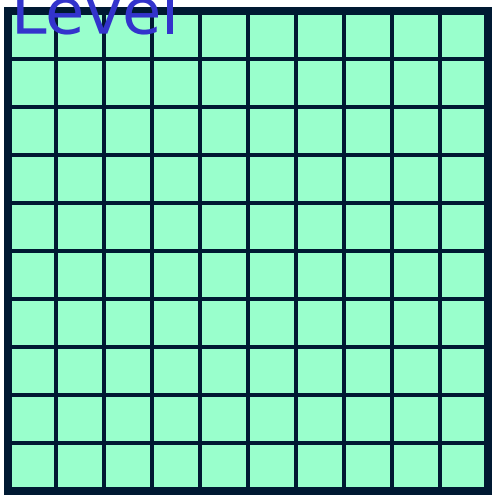
SU about target center (0.5 acre)

If mean of SU < L, conclude DU is "clean."

<u>Key Assumption</u>: SU is the most contaminated portion of DU.

Contamination Equally Probable in DU

L = Action Level



DU = 500 acres

N = 1000 0.5-acre strata

(Only 100 shown above.)

Randomly sample a subset of the strata to determine if DU is "clean."

Sample 57 strata. If $\mu_{0.5} < L$ for all 57 strata (SUs), 95% of DU is clean with 95% confidence.

If N = 10,000, sample 59 strata (< 1% of DU) to demonstrate the DU is "clean."

$$\alpha = P(k,n,M,N) = \frac{\frac{m!}{(m-k)!k!} \times \frac{(N-M)!}{(n-k)!(N-M-n+k)!}}{\frac{N!}{n!(N-n)!}}$$

Hypergeometric Probability Distribution

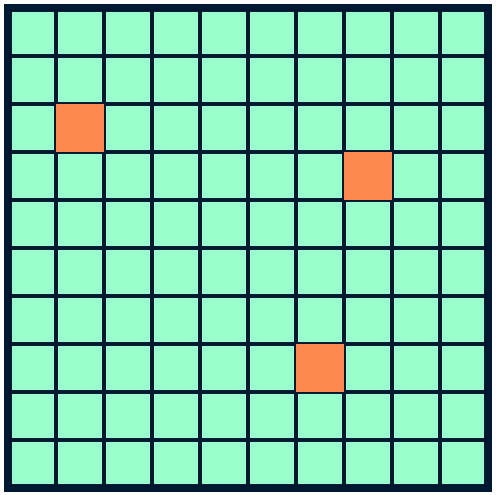
P(k, n, M, N) = Probability of k "dirty" strata (μ > L) out of n sampled strata given a total of N strata where M are "dirty."

p = proportion of "clean" strata in DU (e.g., p = 0.95)
1 - α = Level of confidence (e.g., α = 0.05).

Substitute α, k = 0, N, $M = N(1 - p)$ and solve for n.

UXO module in Visual Sampling Plan (VSP) calculates n.

What if only 3 analyses are funded?



DU = 500 Acres

N = 1000 0.5-acre strata

Randomly sample 3 strata.

If $\mu_{0.5}$ < L for each SU, there is

80% confidence 50% of DU is clean OR

50% confidence 75% of DU is clean.

- Probabilistic methods are NOT inherently more costly than judgmental designs; both can be done using small sample sizes (*n*). However, unlike probabilistic methods, judgmental methods don't report the large decision errors that would result.

What is a sample?
[statistical interpretation]

1. State problem (objective):

   Show that site soils have acceptable levels of lead (Pb).

2. Identify Decision:

   Determine if mean Pb conc. in soils < 400 mg/kg.

   [Determine if null hypothesis H_0: $\mu \geq 400$ can be rejected]

3. Identify decision inputs:

   Conc. of Pb in 1-g aliquots of soil (from grab or composite) analyzed by Method 6010B

   [*X*, variable]

4. Define study boundaries:

DU = SU = Firing range - Soil enclosed in a 10 m × 10 m × 1 m (100 m^3) volume. [Population]

Population size, N = Number of 1-g aliquots in population

N = Volume of soil × Density of soil × Aliquots per unit mass

≈ 10^8 cm^3 × 1.5 g/cm^3 × 1 aliquot/g

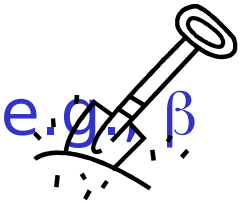
= 1.5 × 10^8 (150 million) aliquots

5. Develop decision rule: Conclude $\mu < 400$ if upper confidence limit (UCL) of mean (e.g., 95% UCL) < 400

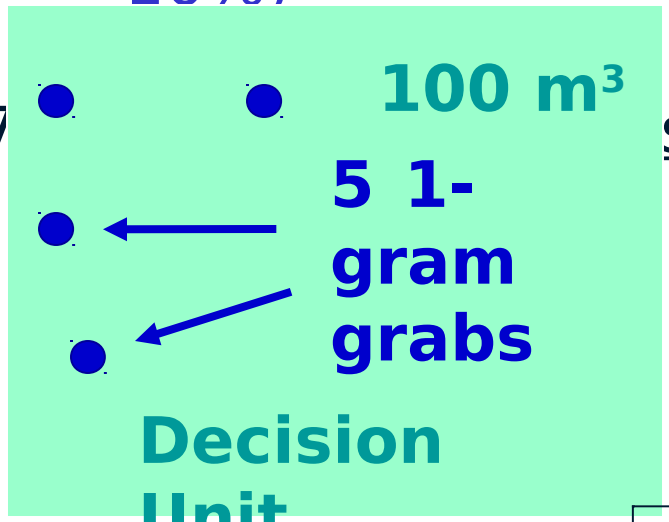
6. Develop tolerances for decision errors:

   Type I error (α) – Wrongly conclude $\mu < 400$ (e.g., α = 5%)

   Type II error (β) – Wrongly conclude $\mu \geq 400$ (e.g., β = 10%)



7. Design sampling design.



How many samples were taken?

One. Set of n = 5 grabs is **the sample** to statistician; each grab is a "specimen." (Step 6 determines n.)

Sample

- **Mathematically**, a set of n measurements of a variable X

$\{x_1, x_2, x_3, \ldots x_n\}$, where n = sample size

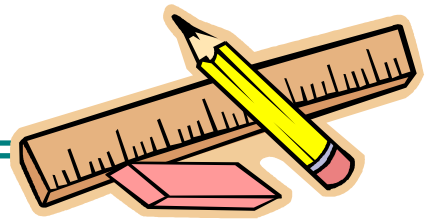
to estimate population (SU) parameters

Example: Sample mean (sample statistic)

$$\bar{x} = (x_1 + x_2 + x_3 + \ldots + x_n) / n \quad (n = 5)$$

estimates population mean (population parameter)

$$\mu = (x_1 + x_2 + x_3 + 2\Omega + x_N) / N \quad (N = \times 10^8)$$



- **Data evaluation: If data are normal:**

$$95\%\ \text{UCL} = \bar{x} + t_{0.95,\, n-1}\, s / n^{1/2}$$

s = sample standard deviation, $\bar{x}$ = sample mean

t = 95th percentile Student's t distribution, $df = n - 1$

If UCL < 400, conclude μ < 400 for the SU with at least 95% confidence.

Key Underlying Assumption

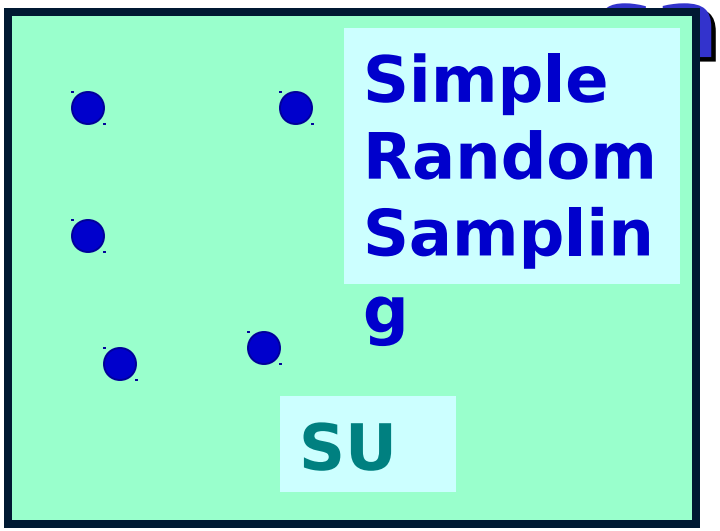
- Sample physically **represents** the entire SU unit, which largely depends on sampling design. Otherwise, the statistical evaluations will not be meaningful.
- Sample will not be representative of SU if biased or judgmental (grab or composite) sampling is done. It will not be possible to make reliable inferences (decisions) from the sample.

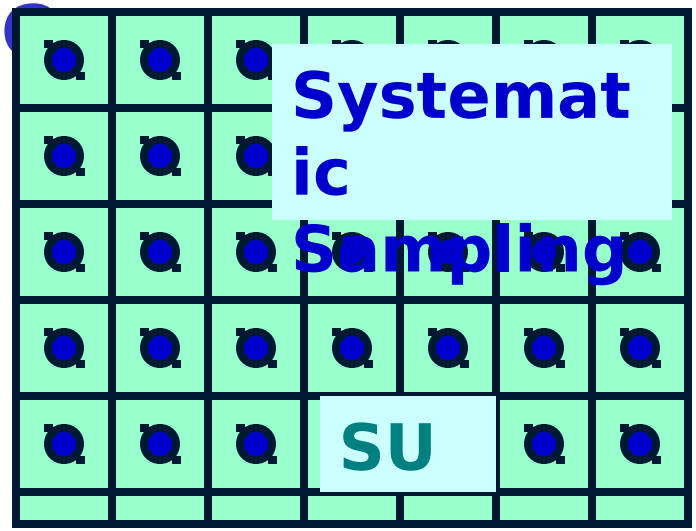
Biased Sampling

EPA QA/G-5S

Judgmental sampling does **not** allow the level of confidence (uncertainty)…to be accurately quantified. In addition, **judgmental sampling** limits the inferences made to the units actually analyzed, and extrapolation from those units [the sample] to the overall population [SU]…is subject to **unknown** selection **bias**…

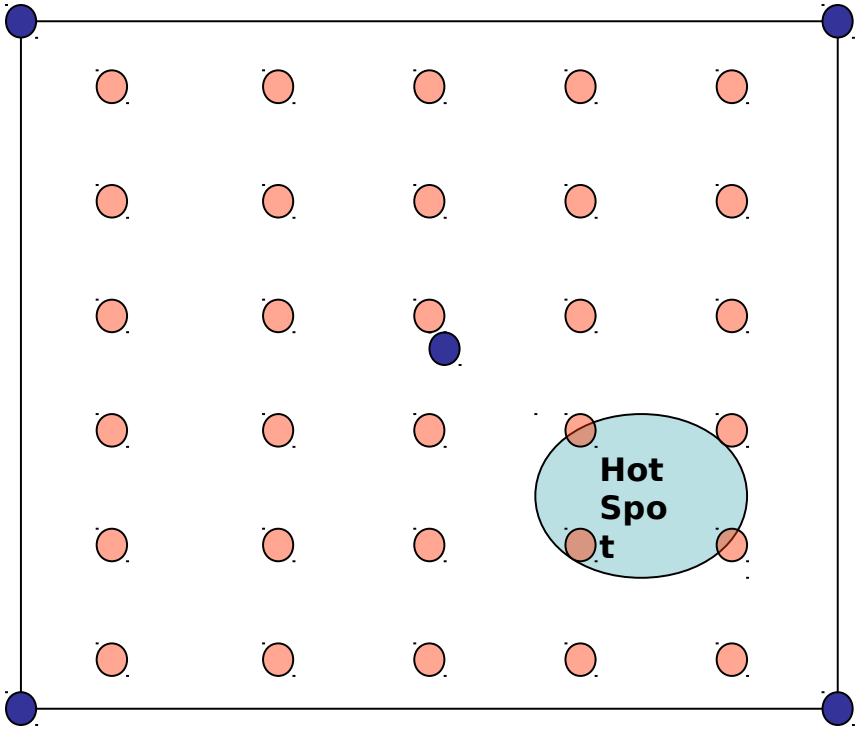
Collecting a representative sample





- A statistically-based sampling design is needed (e.g., per EPA QA/G-5S guidance).
  - **Random** – Every sample (set of *n* measurements) has an equal opportunity of being selected.

- Sample will not be representative if "coverage" of SU is poor (e.g., because sample size or density is too small).
- Heterogeneous environmental media (e.g., soils & sediments) require composites or large numbers of grabs.



Grabs (blue dots) "miss" "hot spot" but contamination owing to "hot spot" is captured if a 30-increment composite (pink dots) or equal number of grabs are collected.

True of False? It is "conservative" to compare the sample maximum (max) to a decision limit using a small number of grabs.

- Action Level for DNT in soil = 0.75 ppm

- *n* = 5 grabs:

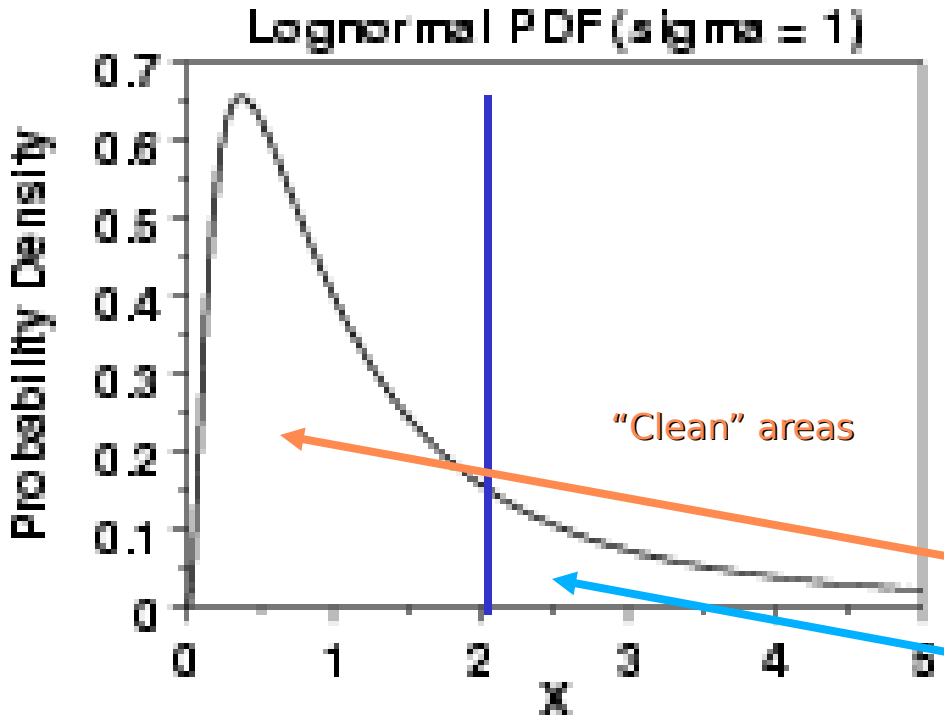
  0.51, **0.61 (max)**, 0.12, 0.16, 0.39 ppm

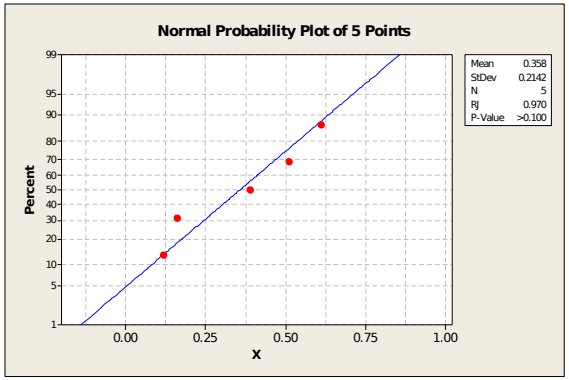
- Conclusion:

  SU is "clean" as **0.61 ppm < 0.75 ppm**.

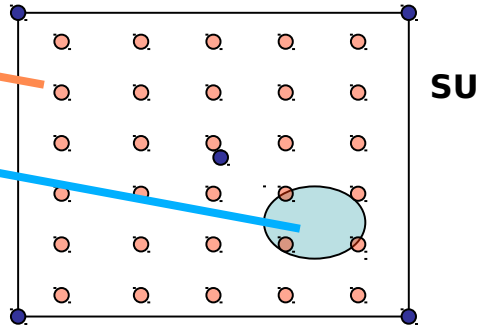
- 5 points randomly taken from lognormal distribution with $\sigma = 2$ and $\mu = 1.1$ **> 0.75**

Environmental distributions tend to be positively skewed. If *n* is small, detection of extreme values will be unlikely. Variability will seem small and the distribution normal.





Right tail: Extreme value from "hot spot."



Comparing Sample Maxima to Decision Limits Using Grabs

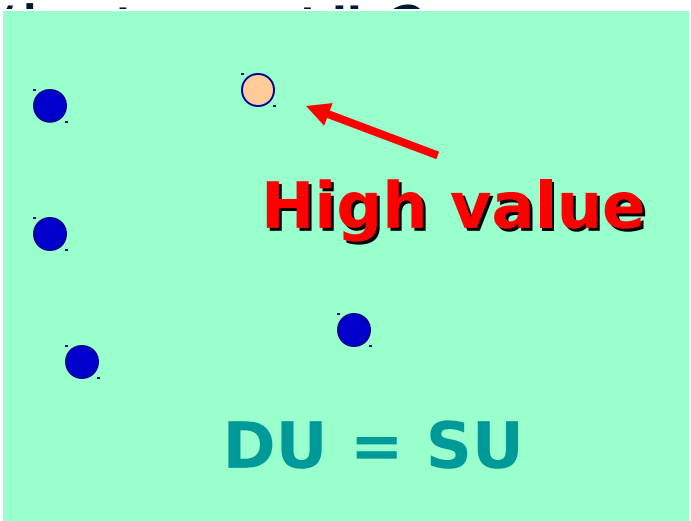
- Distribution cannot be characterized when *n* is small (e.g., *n* needs to be ≥ 15 – 20).
- Reliable inferences cannot be made from the sample max.

  Sample max ≠ Population (SU) max

  SU mean, μ/< sample max (for small *n*).

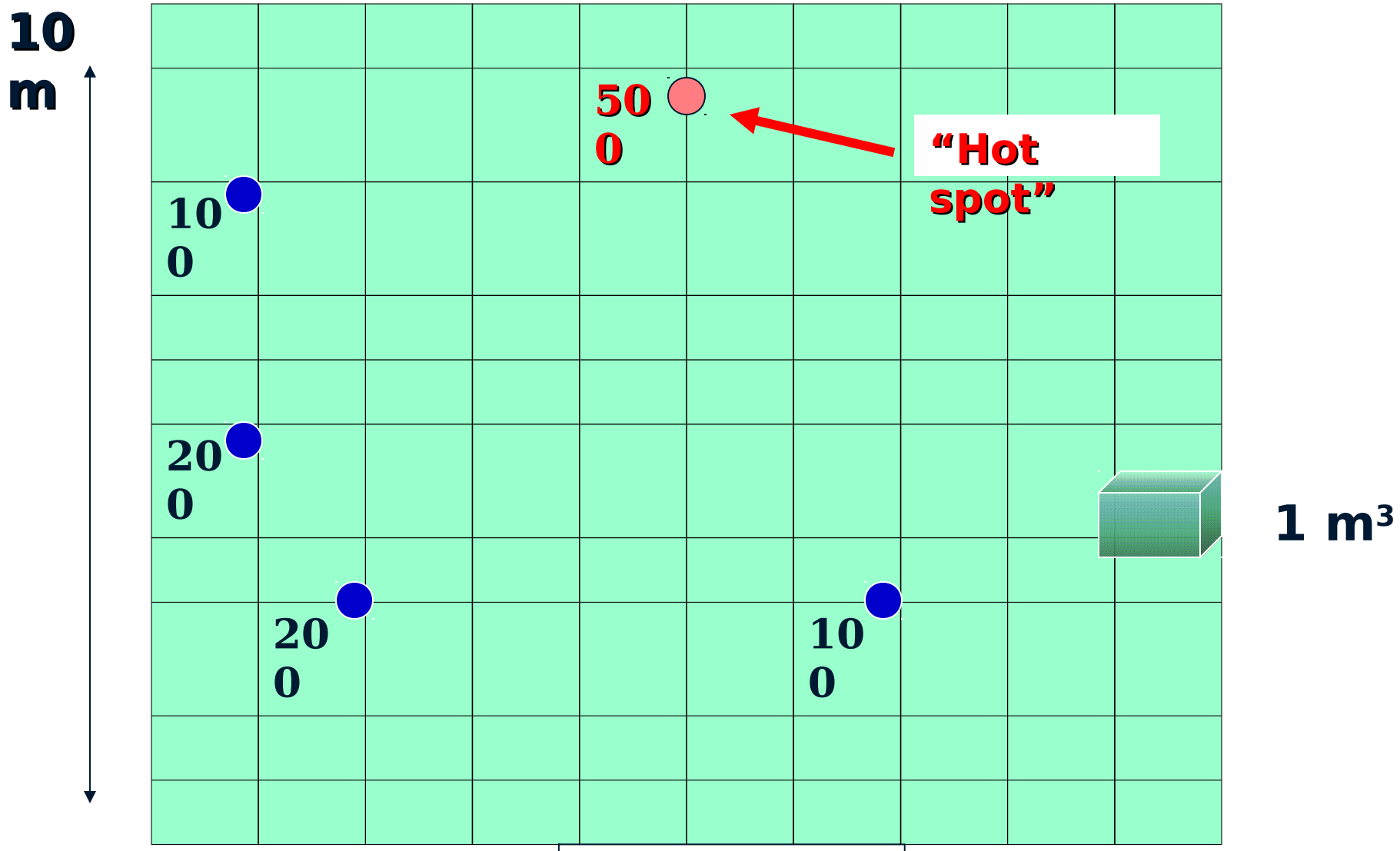
Hot Spots Paranoia

- Five grabs are randomly collected but one exceeds the decision limit (400 ppm). Is the location at which the grab was collected a "hot spot"?

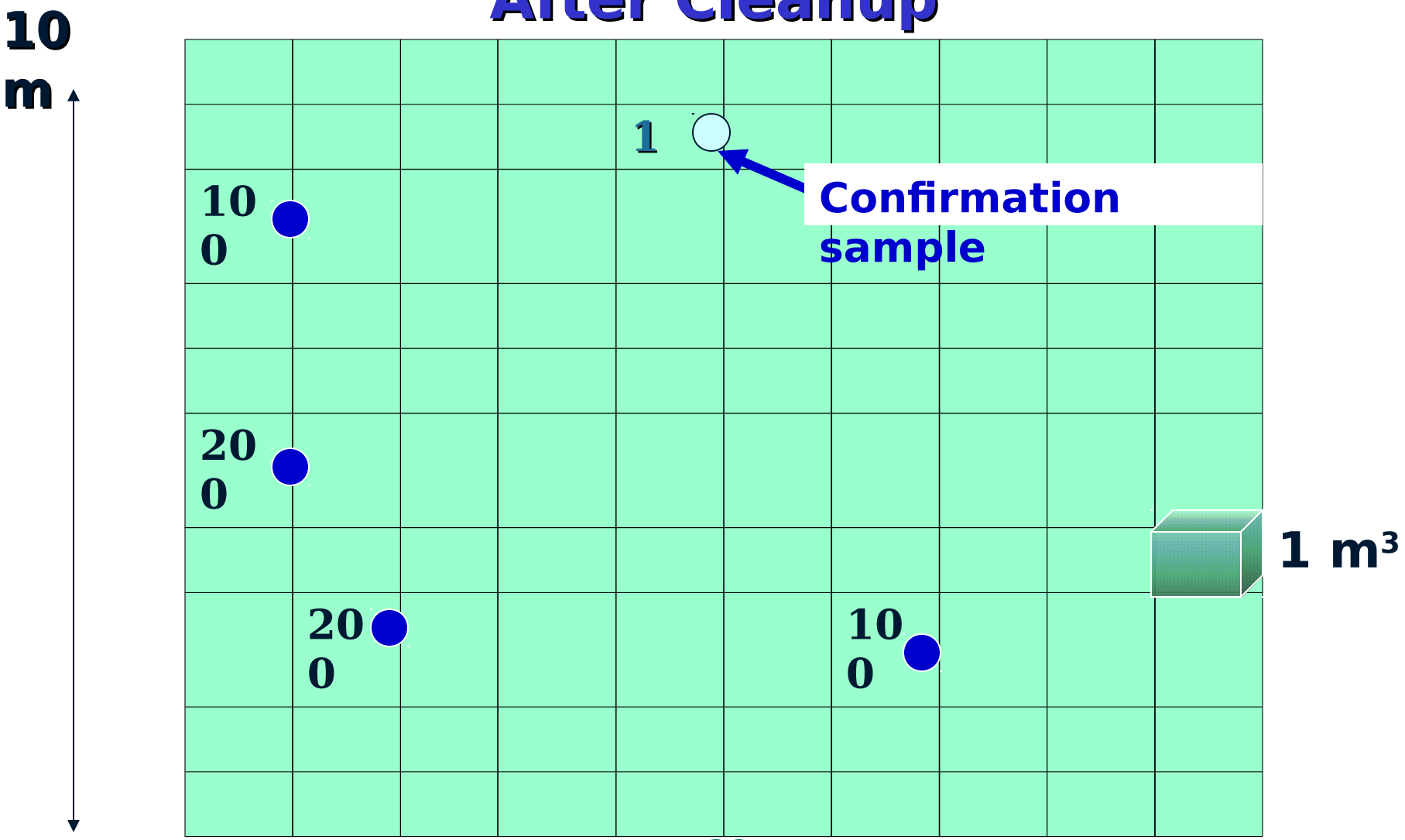


The **set** of grabs is the **sample**; an individual grab does **NOT** represent anything (i.e., some volume of soil). High value likely owing to random variability.

Action Level, L = 400



After Cleanup



Before Cleanup: μ = 266, 33% of site >

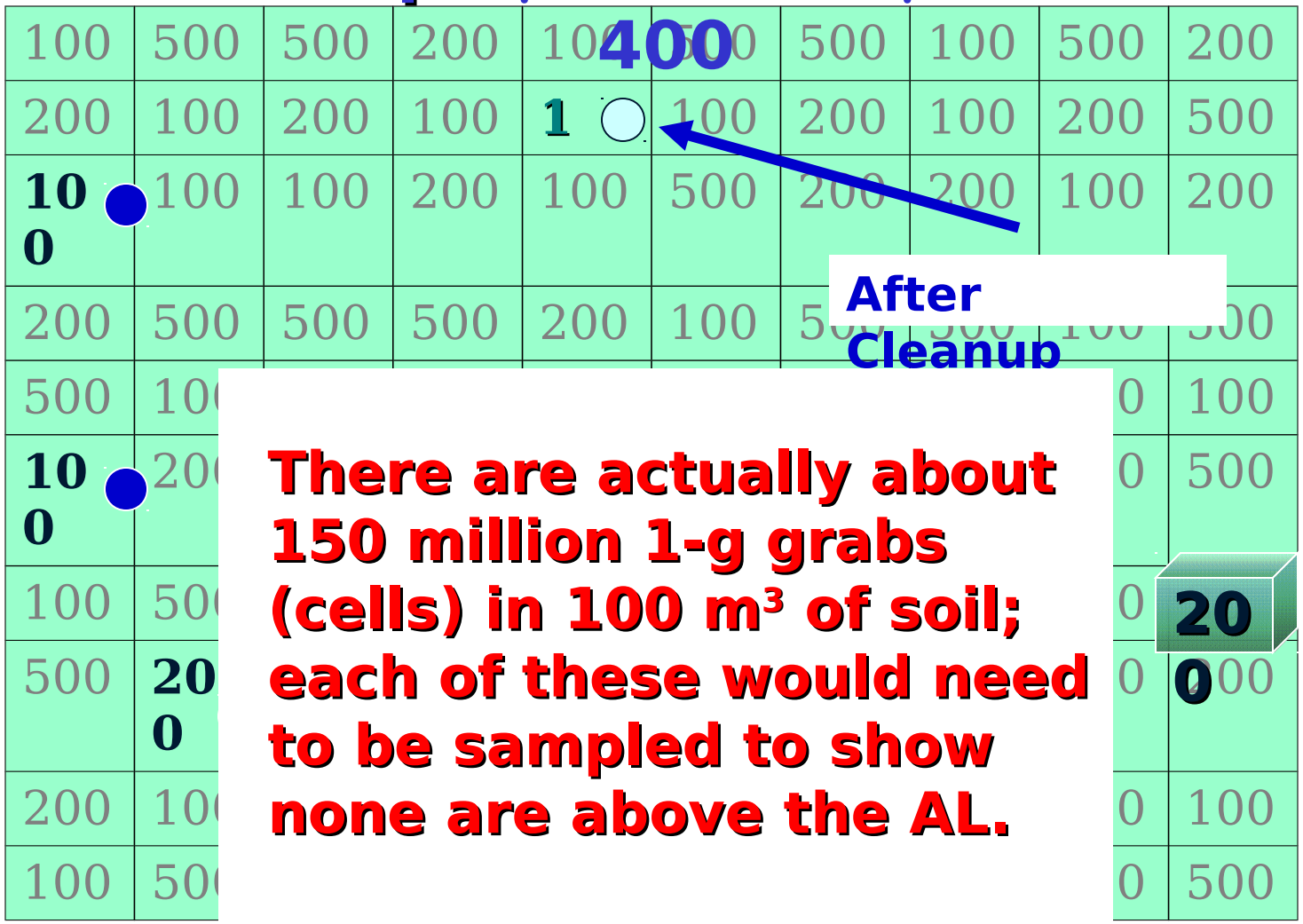
10 m

100	500	500	200	100	500	500	100	100	100
200	100	200	100	500	100	200	100	200	500
100	100	100	200	100	500	200	200	100	200
200	500	500	500	200	100	500	500	100	500
500	100	200	100	100	200	500	500	500	100
100	200	500	500	200	200	100	100	200	500
100	500	100	500	500	200	500	200	100	200
500	200	200	200	100	500	100	200	200	200

"Hot spot"

200 1 m³

After Cleanup: $\mu = 266 \rightarrow \mu = 261 < L =$ 400



After Cleanup

There are actually about 150 million 1-g grabs (cells) in 100 m³ of soil; each of these would need to be sampled to show none are above the AL.

200

1 m³

Defining Hot Spots

Mass (M) = Concentration (C) × Volume (V)

- Any definition of a hot spot that does not address the **C (conc.) & V** (size) of the hot spot relative to the DU (e.g., exposure unit) is not scientifically defensible.

Most definitions of hot spots are incomplete or invalid ! *

- Oregon:

  Any conc. 10 – 100 × acceptable human or ecological risk.

- Washington:

  …small patches of higher level contamination…

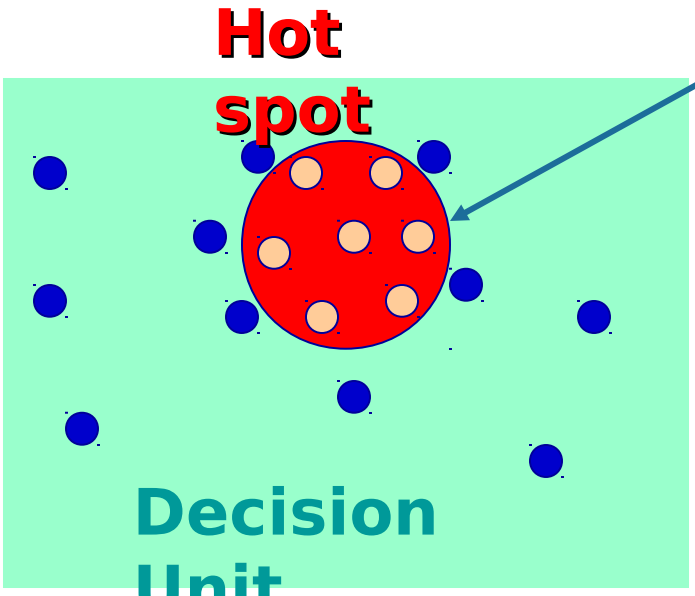
- New Hampshire

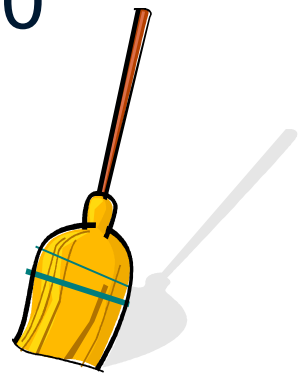
  …any region that exceeds some threshold concentration…

* *Use of Risk Assessment in Management of Contaminated Sites*, Interstate Technology & Regulatory Council, 03/08.

Is cleanup required for "hot spot"?

Action level, L = 10 ppm, $\bar{x}_{\text{Hot Spot}}$ = 100 ppm



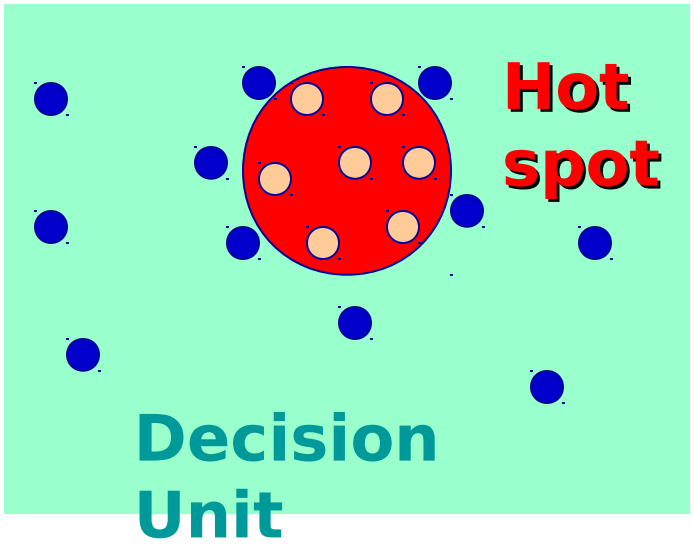


Not necessarily!

Need to know (i) conc. in "clean" area (ii) size of hot spot relative to that of study area.

Calculating a Weighted Mear

$$\bar{x}_{DU} = w_{Hot\ Spot}\,\bar{x}_{Hot\ Spot} + w_{Clean}\,\bar{x}_{Clean}$$



$\bar{x}_{DU}$ = Sample mean of DU

w = Weighting factor

$w_{Hot\ Spot} + w_{Clean} = 1$

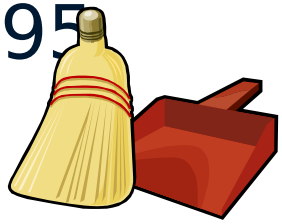
$w_{Hot\ Spot}$ = "Hot spot" area / Total site area

w_{Clean} = "Clean" area / Total site area

Determining if cleanup is needed

$L = 10$, $\bar{x}_{Hot\ Spot} = 100$, $x_{clean} = 1$,

$w_{Hot\ Spot} = 0.05$, $w_{Clean} = 1 - w_{Hot\ Spot} = 0.95$

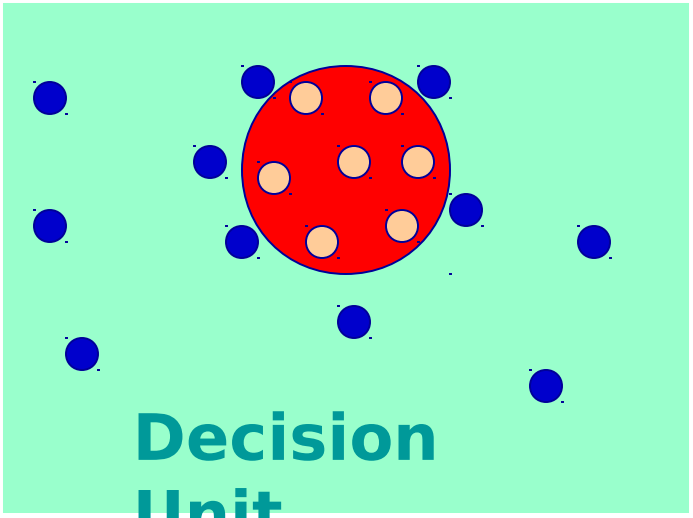


$$\bar{x}_{Site} = w_{Hot\ Spot}\,\bar{x}_{Hot\ Spot} + w_{Clean}\,\bar{x}_{clean}$$

$$= (0.05 \times 100\ ppm) + (0.95 \times 1\ ppm)$$

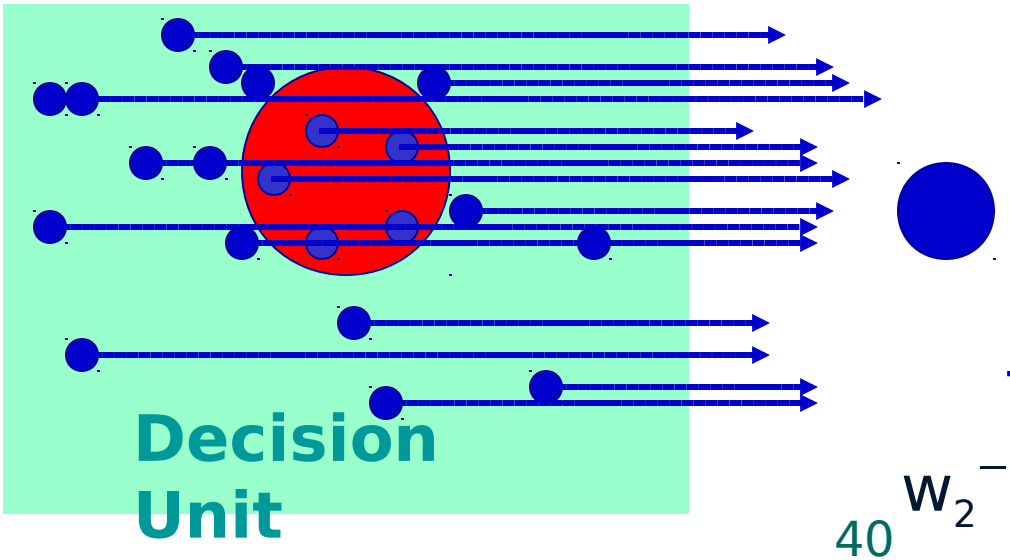
$$= \mathbf{5.95\ ppm < 10\ ppm}$$

Cleanup not required! "Hot spot" too small for unacceptable risk over entire exposure unit (DU).



DU stratification into 2 SUs with grab sampling.

$$\bar{x} = w_1 \bar{x}_1 + w_2 \bar{x}_2$$



DU = 1 SU with MI Sampling

MIC $\rightarrow \bar{x}_{MIC}$

$$\bar{x}_{MIC} \approx w_1 \bar{x}_1 + w_2 \bar{x}_2$$

Summary of Key Points

See, in my line of work you got to keep repeating things over and over and over again for the truth to sink in, to kind of catapult the propaganda.

- George W. Bush

I know what I believe. I will continue to articulate what I believe and what I believe – I believe what I believe is right.

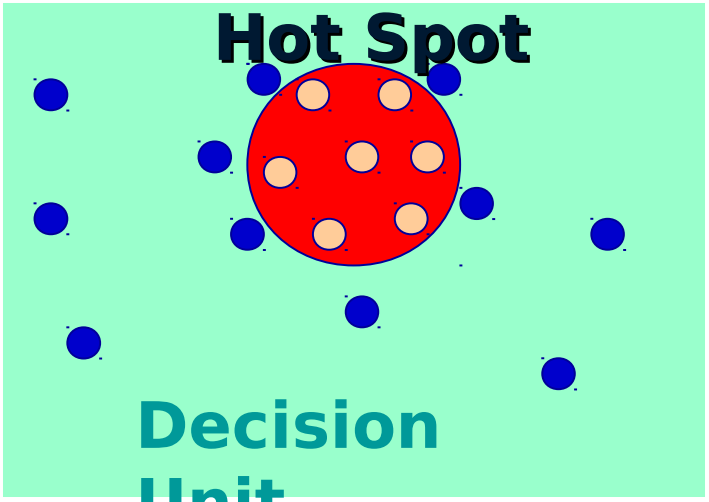
- George W. Bush

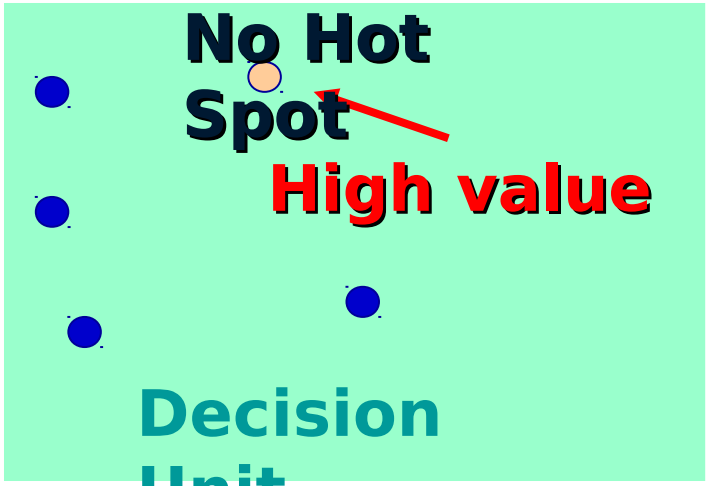
Key Points

- Use 7-Step DQO Process (QA/G-4).
  - Judgmental methods are typically not scientifically defensible, producing errors of unknown magnitude.
  - Define DU & SUs in Step 4 of DQO Process based on end data uses, not necessarily on extent of contamination.
  - **Sample quality ≠ Decision quality!** Collecting a "high quality" sample for a SU (e.g., MI sampling) does not imply 7-Step DQO process is unnecessary.

Hot Spots

- Hot spot ≠ Point. Sample size must be large enough to determine its **C & V** (boundaries).





- Hot spot ≠ DU. It is a subset of a DU of **C & V** large enough to influence a decision for the DU.

Key Points (continued)

- Use statistical sampling designs.
  - No biased sampling (without sound technical rationale).
  - Stratify the DU into SUs using historical information to optimize sampling design.
  - Collect composites or large number of grabs (e.g., $n \geq 15 – 20$) to characterize SUs & underlying distributions.
  - Sample = Set of measurements the SU – Do NOT base decisions on any single measurement.

Questions

- **Contact Information**

402-697-2567

thomas.georgian@usace.army.mil

Monday – Thursday, 6:00 a.m. – 4:30 p.m.